Circular Dichroic Constrained Structure Optimization of Homoalanine Peptides

BERND MAYER,[1] GIANCARLO MARCONI[2]

[1]*Institute for Theoretical Chemistry and Radiation Chemistry, University of Vienna, UZAII, Althanstrasse 14, A-1090 Vienna, Austria*
[2]*Istituto di Fotochimica e Radiazioni, d'Alta Energia, CNR, Via P. Gobetti 101, 40129 Bologna, Italy*

Received 25 June 1999; accepted 18 October 1999

ABSTRACT: This article describes a new method for peptide structure optimization within a Monte Carlo Simulated Annealing (MCSA) framework, namely the cominimization of potential energies under the constraint of a calculated Circular Dichroism (CD) spectrum. We compute potential energy as well as the CD spectrum of every structure generated within the course of the MCSA run, and cooptimize the mean deviation of this calculated spectrum to a corresponding experimental spectrum together with the potential energies of the respective structures using a modified Metropolis Criterion within the MCSA scheme. We compare the performance of this technique with two other MCSA optimization variants—first, a cominimization of potential energies and free energies of solvation; and second, the standard minimization of potential energy alone. We use homoalanines in lengths of 10 and 15, whose optimized structures are highly α-helical, and correspondingly give α-helix characteristic CD signals as the test peptides. This circular dichroic constrained optimization of potential energies is, compared to the other methods, highly efficient in locating α-helical structures with lowest potential energies in both the 10 alanine and the 15 alanine peptide system, within short MCSA runs. The overall structural information embedded in the CD spectrum efficiently guides the optimization towards the native peptide structure. J Comput Chem 21: 270–281, 2000

Keywords: peptide folding; multiple minima problem; circular dichroism; Monte Carlo dynamics; multiple structure optimization

Correspondence to: B. Mayer; e-mail: bernd@asterix.msp.univie.ac.at

Contract/grant sponsor: Austrian Academy of Sciences within APART (Austrian Programme for Advanced Science and Technology)

Contract/grant sponsor: Hochschuljubiläumsstiftung der Stadt Wien

Introduction

One of the major obstacles in protein folding is the postulated NP completeness of the optimization problem.[1] The energy hypersurface, defining the structural space with its large number of degrees of freedom, is highly rugged and complex. The computational performance for optimization of structures based on such prerequisites is, therefore, nonpolynomial in time, which, per definition, results in poor optimization performance.

A variety of computational methods was introduced over the last years, deterministic and stochastic in nature, to extract structural elements of peptides and proteins from scratch, i.e., only based on their sequence. Examples include deterministic routines as constrained molecular dynamics, which has been applied successfully in folding small proteins from the extended to the native conformation,[2] and a variety of stochastic methods, which are usually more efficient in sampling the conformational space of polypeptides. The Monte Carlo (MC) method including the Metropolis Criterion,[3] generates a canonical ensemble of structures, among which local, or even the global minimum may be found. An acceptance probability of stochastically generated structures is computed given by the Boltzmann factor. The efficiency of MC optimization may be increased via full energy minimization at every MC step,[4] as only local minima are compared via the Metropolis Criterion. Naturally, this Monte Carlo minimization method results in additional loss of computational performance, as the full energy minimization is comparably slow.

Monte Carlo Simulated Annealing (MSCA) performs temperature-dependent MC, and is widely applied in structure determination, from host–guest systems[5] to general polymer folding.[6–8] Other methods include Gaussian Density Annealing,[9] Systematic Step Size Variation,[10] and various types of Genetic Algorithms[11] including conformational space annealing.[12] Monte Carlo simulations performed in a multicanonical ensemble, multicanonical algorithms,[13] are, in particular, efficient in optimizing peptide structures, and they, furthermore, exhibit the advantage of calculating thermodynamic properties as free energy differences between random coil and native state.[13]

This article presents a different optimization method, the circular dichroic constrained energy minimization within a Monte Carlo framework, and compares the efficiency of this technique to standard MC (minimization of potential energies) as well as to a MC with modified Metropolis Criterion, which cominimizes potential energies and free energies of solvation.[7] Circular Dichroism (CD) is a measure of the differential absorbance of optically active molecules, for example, peptides, between left and right circularly polarized light. Holzwarth et al.[14] were the first to measure the CD spectrum of an α-helix, which shows a typical double minimum at around 222 and 208 nm, and a maximum at around 190 nm,[15] whereas a random coil generally shows a strong negative band near 200 nm and a weak, positive, or negative band at around 220 nm.[16] The shape of the CD spectrum of peptides is, therefore, directly related to its secondary structure. This structural information in CD spectra is used in a variety of peptide and protein structure prediction methods, which analyze the CD spectrum of a polymer with unknown structure via a linear combination of CD spectra of proteins of known secondary structure composition.[17]

We use a modified method of the theory for the calculation of the optical activity of polymers developed by Tinoco[18] to explicitly calculate the CD spectrum of every peptide structure generated in a MCSA optimization. The acceptance probability of a structure, computed via an extended Metropolis Criterion in the MCSA run, depends not only on the potential energy of the structure, but in addition on the deviation of the computed CD spectrum of this conformation when compared to the respective experimental spectrum of the native peptide structure.

We use homoalanine peptides in lengths of 10 and 15 to test the performance of our different optimization methods. Alanine has a high intrinsic α-helix formation propensity,[19, 20] mainly driven by conformational entropy considerations regarding the side chain.[21, 22] The α-helix formation of short, alanine-based peptides was studied exhaustively using a variety of theoretical methods,[7, 11, 19, 21, 23–26] and also confirmed experimentally, for example, via Circular Dichroism spectroscopy.[27]

The next section introduces the computational details on calculating the Circular Dichroism of peptides as well as on the implementation of the three different optimization variants used, i.e., minimization of potential energy, cominimization of potential energy and free energy of solvation, and finally cominimization of potential energy and deviation of computed and experimentally determined Circular Dichroism.

Computational Method

We apply standard MCSA and MCSA with modified Metropolis Criterion[3, 6] as our optimization framework, i.e., the acceptance of a stochastically generated new peptide structure in a sequence of conformations is given by the Boltzmann probability. According to the classical MCSA scheme, the acceptance probability $P(\Delta E_P)$ of a newly generated structure with the potential energy difference ΔE_P compared to the respective reference structure is given as:

$$P(\Delta E) = e^{(-\Delta E_P/RT)} \tag{1}$$

where R is the gas constant (kcal mol^{-1} K), and T is the temperature in Kelvin.

In the present setup, the potential energy E_P is calculated applying the empirical conformational energy program ECEPP/3.[28] Single-point energies are computed at every MCSA step, and no additional gradient minimization is performed. A multiple exponential cooling scheme is used as the temperature function in the simulated annealing process.[7]

To account for solvation effects we use a continuum approximation[7] to calculate changes in the free energy of solvation, ΔG_{solv}. The following equation gives a modified Metropolis criterion,[5, 7, 29] which leads to a cominimization of potential energies computed by the force field and the free energy of solvation calculated via the continuum approximation, again comparing actual and reference structure.

$$P(\Delta E_P, \Delta G_{\text{solv}}) = e^{(-\Delta E_P/RT)} \times e^{(-w\Delta G_{\text{solv}}/RT)} \tag{2}$$

where ΔG_{solv} is the change in free energy of solvation, and w is the weighting factor.

The acceptance probability $P(\Delta E_P, \Delta G_{\text{solv}})$ reflects both changes in potential energy, ΔE_P, as well as changes in the free energy of solvation, ΔG_{solv}, as the product of both probabilities. To account for the different relative scaling of force field and free energies of solvation a weighting factor is applied. A detailed analysis of the weighting factor for peptide structure optimization is given in ref. 7, and for host–guest systems in ref. 29.

The free energy of solvation, ΔG_{solv}, factorized into contributions assigned to hydrophilic and hydrophobic surface areas, is calculated as:

$$G_{\text{solv}} = \left(\sum_{i=1} \sigma_k A_i\right)_{\sigma_k<0} + \left(\sum_{i=1} \sigma_k A_i\right)_{\sigma_k>0} \tag{3}$$

where A_i is the solvent accessible surface area of atom i, and σ_k is the atomic solvation parameter corresponding to A_i.

First, the solvent accessible surface A_i is computed,[30] and then multiplied with the respective atomic solvation parameter (ASP) σ_k. The ASPs represent the free energy of solvation of group i normalized on the solvent accessible surface area, and are determined experimentally from thermodynamic data for the transfer of model compounds from vapor phase to water.[31] Values of $\sigma_k > 0$ correspond to hydrophobic (carbon), $\sigma_k < 0$ to hydrophilic surface areas (nitrogen, oxygen). The atomic solvation parameter set of Wesson and Eisenberg in the adjustment of Sharp et al.[31] is used in our calculations.

The solvent accessible surface is calculated within MSEED.[30] A probe radius of 1.4 Å is chosen, corresponding to the probe radius used in the evaluation of the atomic solvation parameters.[31]

The fitness function given in eq. (2) represents the first type of constrained minimization discussed in this article, i.e., cominimization of potential as well as of solvation energies. The next paragraph introduces the constrained minimization of potential energy as well as of the deviation of calculated and experimentally determined Circular Dichroism spectrum of peptide structures generated within the MCSA scheme.

Calculation of the Circular Dichroism of Peptides

The basic theory for the calculation of the optical activity of a polymer has been developed by Tinoco[18] and widely applied and reviewed by Woody.[32] The rotational strength arising from a set of groups embedded in the static field of a polymer is given by three contributions:

$$R_{0a} = \sum_{i=1}^{N} \Bigg[Im\, \mu_{i0a} m_{i0a} - 2 \sum_{j\neq i} \sum_{b\neq a} \frac{Im\, V_{i0a,j0b} (\mu_{i0a} m_{jb0} \nu_a + \mu_{j0b} m_{ia0} \nu_b)}{h(\nu_b^2 - \nu_a^2)} - \frac{2\pi}{c} \sum_{j\neq i} \sum_{b\neq a} \frac{V_{i0a,j0b} (R_j - R_i)(\mu_{j0b} \times \mu_{i0a}) \nu_a \nu_b}{h(\nu_b^2 - \nu_a^2)} \Bigg] \tag{4}$$

where N gives the total number of residues, Im indicates the imaginary part, μ gives the electric and m the magnetic transition moments, R_i and R_j are the vector distances between the centers of the ith and

jth residue, and ν_i gives the frequency of the transition $0 \rightarrow I$.

V represents the geometrical factor, expressed in tensorial form as:

$$V_{i0a,j0b} = \boldsymbol{\mu}_{j0a} \boldsymbol{T}_{ij} \boldsymbol{\mu}_{j0b} \tag{5}$$

$$\boldsymbol{T}_{i,j} = 1 - \left(\frac{3\boldsymbol{R}_{ij}\boldsymbol{R}_{ij}}{\boldsymbol{R}_{ij}^2} \right) \tag{6}$$

The presence of all three contributions, expressed in eq. (4), corresponds to different mechanisms that are able to generate optical power in polypeptides: the first mechanism depends on the product of the electric and magnetic moments of the residues, and gives zero for achiral groups. The second contribution arises from the coupling of electric and magnetic moments located on different residues, modulated by the geometry factor and by the different frequency of the transitions involved. This term, which was neglected in many previous treatments of the problem, as discussed in ref. 33, is known to play an important role especially for groups having low-lying $n\pi^*$ states interacting with higher energy $\pi\pi^*$ states of other groups.[34] The third term finally depends on the interaction of electric dipole moments in different groups, and is known to generate the most important contribution considering allowed transitions. In case of identical chromophores, this term is known as Moffits exciton term.

The calculation of the contribution to R_{0a} of a number of N identical residues for each band requires the solution of the secular equation:

$$\sum_{i=1}^{N} C_{iak} V_{i0a,j0a} - E_{AK} C_{jaK} = 0 \tag{7}$$

$$\nu_{AK} = \nu_a + \frac{E_{AK}}{h}$$

The molecular properties of the residues can be calculated in principle by quantum mechanical methods, or preferably, taken by the spectra of the compounds. This method is adopted in the present case for the electric moments and energies of the two lowest excited states of alanine.[35] However, the magnetic properties and the intrinsic rotational strength of the residues are calculated quantum mechanically within the method of the complete angular momentum operator[36] adopting CNDO/S wave functions.

The final circular dichroism spectrum is obtained by a sum of Gaussians centered on the calculated wavelengths of the various bands along the expression:

$$[\vartheta_i^0] = a\left(\frac{\lambda_i}{\Theta_i}\right) R_{0i} \tag{8}$$

$$[\vartheta_i] = [\vartheta_i^0] e^{[-(\lambda_i - \lambda)^2/\Theta_i^2]}$$

where Θ_i is the half-width of the band centered at the wavelength λ_i, and a is a constant amounting to 0.811×10 exp. 42, where R_{0i} is expressed in CGS units.[35]

Figure 1 shows the computed rotatory strength R_{0a} for homoalanine α-helices in lengths of 5, 10, and 15, plotted vs. the wavelength given in nanometers.

The analysis of the CD spectrum of an α-helix polypeptide shows that the lowest wavelength contribution (negative at 222 nm) arises from the $n\pi^*$ amide transition, whereas the bands at 202 and 190 nm (negative and positive, respectively) derive essentially from the exciton splitting of the $\pi\pi^*$ amide transition.[32] The calculated spectrum shows that the negative band at around 222 nm as well as the positive band at around 190 nm are in good agreement with the experiment,[27, 37] whereas the experimentally found negative band at around 202 nm is less well reproduced for homoalanine. It is found that diagonalization of eq. (7) results in a pivotal concentration of rotational strength on the positive, short wavelength side of the spectrum with respect to the long wavelength side, which appears underestimated. The relative balance of these two bands is dependent on the nature, properties, and geometry of the residues. In fact, preliminary calculations carried out on an alanine-lysine peptide showed a clear negative band around 205 nm, which is in good agreement with the experimental spectrum of this helical peptide.[27] The computation of the CD, fur-

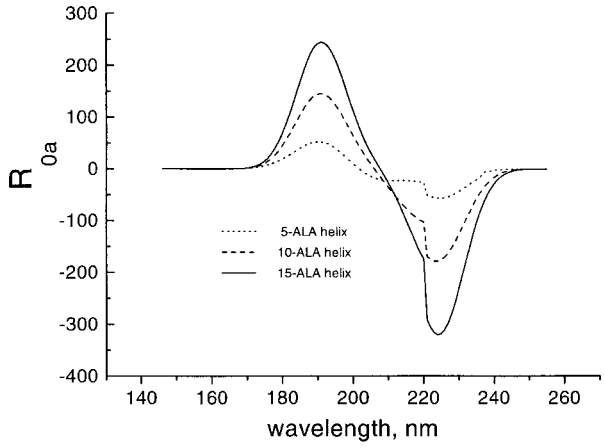

FIGURE 1. Calculated rotatory strength R_{0a} of the α-helical homoalanine peptides in lengths of 5, 10, and 15 plotted vs. the wavelength λ given in nanometers.

thermore, nicely reproduces the relative increase of the rotatory strength when comparing helices with increasing lengths.[38]

Circular Dichroic Constrained Optimization of Peptide Structures

To include the calculated spectrum, as shown in Figure 1, as well as the corresponding experimental spectrum in the MCSA optimization, both are normalized to the interval $[-1, 1]$ and the following expression is computed, giving CD_{err}:

$$CD_{\mathrm{err}} = \frac{\sum_{i=1}^{n} \left[\frac{cd^{\mathrm{exper.}}(i)}{cd^{\mathrm{comput.}}(i)} \right]}{n} \qquad (9)$$

where $cd^{\mathrm{exper.}}(i)$, $cd^{\mathrm{comput.}}(i)$ are the experimentally determined and computed rotatory strength at wavelength i in the natural number interval of interest ($n \in [160, 260]$ in the present case).

ΔCD_{err} finally gives the difference in CD_{err} comparing actual structure and respective reference structure in the course of the MCSA algorithm, and is included in a modified Metropolis criterion as

$$P(\Delta E_P, \Delta CD_{\mathrm{err}}) = e^{(-\Delta E_P / RT)} \times e^{(-w\, \Delta CD_{\mathrm{err}})} \qquad (10)$$

where w is the weighting factor.

This criterion leads to a constrained minimization of potential energy, as it is coupled to an optimization of the calculated CD spectrum with respect to the experimentally determined spectrum. Again, a weighting factor is used to balance the optimization constraint coming from energetic as well as spectral considerations.

Equation (10) is analogous to eq. (2), resulting in a cominimization of potential energy and CD spectrum, as well as cominimization of potential energy and free energy of solvation, respectively. The prerequisite for applying eq. (10) as the fitness function is the, usually accessible, CD spectrum of the peptide of interest.

The following Figure 2 gives a flow chart of the program package MultiMize holding the routines to calculate the potential energy E_P, the free energy of solvation G_{solv}, as well as the CD spectra and the respective CD_{err} within the MCSA optimization scheme.

The peptide optimization test system presented in this article uses homoalanines in lengths of 10 and 15 (10-ALA and 15-ALA), and refers to the spectra shown in Figure 1 as the experimental spectra of the native structures, as homoalanine peptides are not soluble in water and, therefore, their CD is not experimentally accessible. Quadrifoglio et al.[34] report the helix formation of polyalanine in trifluoroethanol, and Baldwin et al.[24] explored the helix formation of closely related peptides with three lysines used to mediate solubility in aqueous buffer, and they proved experimentally the helix formation of such peptides. The high helix formation tendency of homoalanine and alanine-based peptides was, furthermore, confirmed by extended conformational analysis,[19] Monte Carlo methods,[7] Systematic Step Size Variation,[11] as well as molecular dynamics.[21, 23]

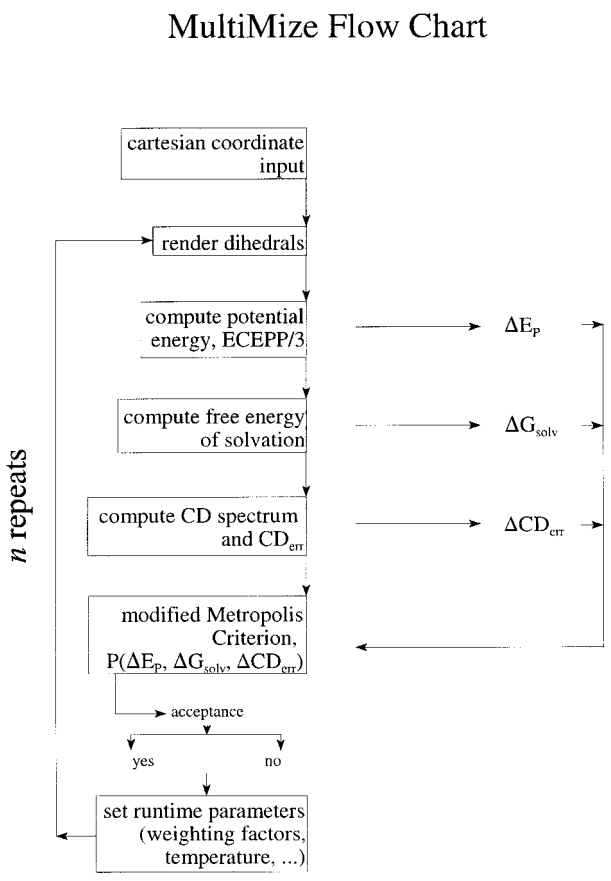

FIGURE 2. Flow chart of constrained optimization within the program package MultiMize. The difference in potential energy, ΔE_P, and the difference in solvation energy, ΔG_{solv}, of the actual structure with respect to the reference structure is computed. The next step determines the mean deviation of the computed CD of the actual structure compared to the experimental spectrum of the native structure, CD_{err}. Again, comparing actual and reference structure gives ΔCD_{err}. The modified Metropolis Criterion finally computes a probability P, which may depend on both energetic contributions as well as on the value of ΔCD_{err}. After adjustment of run-time parameters, as the temperature and weighting factors, the iteration cycle is restarted n times.

Computational Details

For every MCSA step one randomly chosen rotatable bond is stochastically updated in the range $[-180, 180]$. Rotatable bonds are the ϕ, ψ, as well as side-chain dihedrals (ω angles are fixed to 180°). The total number of degrees of freedom is 30 for the 10- and 45 for the 15-residue homoalanine peptide, respectively. The start coordinates for the MCSA runs are extended conformations (all ϕ, ψ, ω and the side-chain dihedrals χ are equal 180°) of 10-ALA and 15-ALA with neutral termini, and 10^5 MCSA steps are computed for each individual optimization run. A multiple exponential cooling scheme is used as the temperature function with a start point of 1000 K and an end point of 278 K, and one sudden heat-up back to 1000 K at MCSA step 5×10^4.[7]

Five runs with different random update schemes are performed for each optimization *method*, i.e., minimizing the potential energy, *pot*, cominimizing potential energy and free energy of solvation with the weighting factors 0.5 and 1.0, *pot_solv0.5* and *pot_solv1.0*, and finally cominimizing potential energy and CD_{err}, with the weighting factors 1.0, 3.0, 5.0, and 10.0 leading to *pot_cd1.0*, *pot_cd3.0*, *pot_cd5.0*, and *pot_cd10.0*.

Peptide structures are characterized by their potential energy E_P, their total solvation energy G_{solv}, and their end: end distance given as the distance between terminal $\mathbf{N}H_2$ and **C**OOH. A set of parameters defines their structure with respect to an ideal α-helix. A residue k is considered "helical" if $\phi_k \in [-60 \pm 45]$ and $\psi_k \in [-50 \pm 45]$. The corresponding parameters are *no.h.d.*, the total number of "helical" residues, *no.cons.h.d.*, the total number of *consecutive* "helical" residues (also denoted as *helix length*), and finally *dihed.err*, giving the mean deviation of all ϕ, ψ pairs from their ideal values $[-60,-50]$.

The calculations are performed on an IBM RS-6000/375. The CPU time comparing the three methods, i.e., *pot*, *pot_solv*, and *pot_icd* scales roughly as 1 : 1.3 : 1.25.

Results and Discussion

The following optimization results refer to the 10-ALA and 15-ALA homopeptide and give energetic and structural data regarding the optimization performance comparing the different methods, i.e., solely minimizing the potential energy, cominimizing potential and solvation energy, as well as cominimizing potential energy and deviation between calculated and experimentally determined Circular Dichroism spectra.

Table I gives energetic and structural parameters describing the fully extended ($\phi = \psi = 180°$), and ideal α-helical conformation ($\phi = -58°$, $\psi = -47°$) of 10-ALA and 15-ALA, respectively.

The helical conformation is the preferred "native" structure of homoalanine peptides, considering the potential energy E_P. However, the change in total solvation energy of the folded structure compared to the extended conformation is due to shielding of atomic surfaces with negative ASPs, positive.[7] This gives, comparing extended and ideal helix, a decrease in E_P of −92.3 kJ/mol for the 10- and −209.6 kJ/mol for the 15-ALA system, and an increase in G_{solv} of +83.1 kJ/mol for 10- and +134.5 kJ/mol for the 15-ALA. The helical structure of 10-ALA is energetically (considering both, potential and solvation energy) much less defined compared to 15-ALA with respect to a random coil situation. Most probably, peptides in a length of 10

TABLE I.
Reference Energies and Conformational Parameters for the 10-ALA and 15-ALA Homopeptide.

System	E_P	G_{solv}	*No. cons. h. d.*	*Dihed. err.*	*End : end dist.*
10-ALA, ext.	119.51	−115.45	0	119.05	34.46
10-ALA, hel.	27.21	−32.35	10	2.73	16.03
15-ALA, ext.	177.44	−154.66	0	123.65	53.42
15-ALA, hel.	−32.20	−20.11	15	2.45	23.43

System denotes the 10-ALA homopeptide and the 15-ALA homopeptide in extended (*ext.*) and helical (*hel.*) conformation. E_P gives the ECEPP/3 conformational energy in kJ/mol, G_{solv} gives the total free energy of solvation in kJ/mol [see eq. (3)], *no.cons.h.d.* gives the number of consecutive backbone ϕ, ψ dihedral pairs within the range defined for an α-helix, *dihed.err.* shows the mean deviation of the backbone ϕ, ψ dihedrals from their ideal α-helix values, the *end : end dist.* gives the distance between terminal $\mathbf{N}H_2$ and **C**OOH in Å.

will not form a defined secondary structure in water, whereas 16-residue, alanine-based peptides are, in contradiction to older studies,[39] capable of forming defined and stable helices.[27] These energetic parameters as well as the CD signals inbetween random coil and native structure (as given in Fig. 1) define the boundaries for our three different peptide optimization strategies discussed above.

Table II summarizes energetic and structural parameters on optimizing the structure of 10-ALA and 15-ALA within 10^5 MCSA steps, starting with an extended conformation ($\phi = \psi = 180°$).

Table II denotes ranges (computed over five individual runs for every method) of potential energy, solvation energy, and parameters defining the secondary structure of the two model compounds 10-ALA and 15-ALA (see Computational Details section). All values refer to the structure with lowest potential energy found within a particular MCSA run. Again, the *pot* method means minimization of potential energy [see eq. (1)], *pot_solv* defines a cominimization of potential and solvation energy with the weighting factors 0.5 and 1.0 [see eq. (2)], and *pot_cd* gives the cominimization of potential energy and deviation of experimental and computed CD spectrum with the weighting factors 1.0, 3.0, 5.0, and 10.0 [see eq. (10)].

Minimizing Potential and Solvation Energy

10^5 MCSA steps are, in general, not sufficient to locate the "native," i.e., helical structure of either a length of 10-homoalanine or of 15-ALA solely minimizing the potential energy, i.e., applying the *pot* method.[7] The helix length is in the range of 5–7 in 10-ALA, but the respective structures are energetically preferred compared to the ideal helix (see Table I). 10^5 MCSA steps are, for a problem with 30 degrees of freedom, sufficient to find good local minima on the potential energy hypersurface, although these local minima are not yet the native, helical structure.

TABLE II.

Comparative Ranges of Minimum Energies in kJ/mol and Conformational Parameters (out of Five Individual MCSA Runs per Method) for the 10-ALA and the 15-ALA Homopeptide Located by the Different Optimization Methods.

System	*Method*	*min. E_P*	G_{solv}	*No. h. d.*	*No. cons. h. d.*	*Dihed. err.*	*End : end dist.*
10-ALA	*pot*	(−4.90)–(−5.71)	(−34.69)–(−39.67)	7–8	5–7	27.94–43.64	13.26–14.07
10-ALA	*pot_solv0.5*	(−11.62)–(25.08)	(−35.74)–(−40.34)	6–8	6–8	26.88–35.93	13.96–15.78
10-ALA	*pot_solv1.0*	89.91–91.58	(−65.71)–(−86.48)	3–4	1–3	56.56–65.17	14.61–15.60
10-ALA	*pot_cd1.0*	(−15.68)–44.56	(−34.57)–(−63.58)	6–9	5–9	15.85–51.44	14.23–15.10
10-ALA	*pot_cd3.0*	*(−2.83)–(−4.31)*	*(−34.61)–(−37.67)*	*8–10*	*8–10*	*13.21–28.57*	*13.40–15.75*
10-ALA	*pot_cd5.0*	4.60–53.13	(−39.54)–(−32.52)	6–8	5–8	19.07–32.97	14.50–14.96
10-ALA	*pot_cd10.0*	91.96–115.70	(−63.49)–(−69.18)	5–5	4–3	50.47–62.30	12.07–15.84
15-ALA	*pot*	61.74–80.92	(−64.79)–(−34.99)	5–7	2–4	55.27–66.87	8.54–12.35
15-ALA	*pot_solv0.5*	45.34–83.22	(−58.48)–(−70.27)	7–8	4–6	45.61–49.73	17.57–18.59
15-ALA	*pot_solv1.0*	137.61–161.52	(−110.81)–(−134.93)	5–6	1–2	55.18–61.98	13.67–15.27
15-ALA	*pot_cd1.0*	113.86–127.53	(−71.98)–(−79.17)	2–5	1–2	66.97–82.22	3.57–9.76
15-ALA	*pot_cd3.0*	*(−8.11)–(−48.24)*	*(−23.16)–(−29.13)*	*12–15*	*11–15*	*12.64–27.46*	*20.49–21.83*
15-ALA	*pot_cd5.0*	(−10.32)–(21.42)	(−31.60)–(−40.21)	13–15	13–15	21.39–26.89	17.62–21.97
15-ALA	*pot_cd10.0*	78.25–127.41	(−49.74)–(−93.63)	5–11	3–9	75.18–33.41	14.07–15.32

System denotes the 10-ALA and 15-ALA homopeptide, respectively. *Method* gives the type of optimization: *pot* denotes minimization of the potential energy only; *pot_solv* denotes cominimization of potential energy and free energy of solvation [see eq. (2)] applying the weighting factors 0.5 (*pot_solv0.5*) and 1.0 (*pot_solv1.0*); *pot_cd* means cominimization of potential energy and the deviation of experimentally determined and computed CD spectrum [see eq. (10)], the weighting factors 1.0 (*pot_cd1.0*), 3.0 (*pot_cd3.0*), 5.0 (*pot_cd5.0*), and 10.0 (*pot_cd10.0*) are applied.

Min. E_P gives the lowest potential energy in kJ/mol of a structure found within a particular MCSA run (all other parameters given next refer to these structures); G_{solv} denotes the free energy of solvation in kJ/mol, *no.h.d.* gives the total number of ϕ, ψ dihedral pairs within the range defined for an α-helix, *no.cons.h.d.* gives the consecutive ϕ, ψ dihedral pairs within the range defined for an α-helix (i.e. the helix length), *dihed.err.* denotes the mean deviation of the backbone ϕ, ψ dihedrals from their ideal α-helix values, the *end : end dist.* gives the distance between the terminal groups **N**H_2 and **C**OOH in Å.

In the 15-ALA system with 45 freely rotatable bonds the situation is worse using this minimization approach, as the overall helix length is in the range 2–4, and the potential energies are far above the respective value given in Table I. In this second case, 10^5 steps are neither sufficient to reach structures close to the native conformation nor to locate good local minima on the energy surface. A previous study on peptides of comparable length showed that at least 2×10^5 steps (including full gradient energy minimization at every MCSA step) are needed to obtain helical, low energy structures.[7] The fact that under this optimization setup neither the 10-ALA nor the 15-ALA peptides are folded to a helix is also reflected by the solvation energies of the respective low energy structures compared to the values for G_{solv} given in Table I.

Including the solvation term with the weight $w = 0.5$ into the structure optimization [see eq. (2)], i.e., using *pot_solv0.5*, increases the helix length slightly to 6–8 for 10-ALA and to 4–6 for 15-ALA, with a consecutive decrease in G_{solv} compared to the results obtained under the *pot* method.

Using $w = 1.0$, i.e., the *pot_solv1.0* method, however, does not result in even longer helices, as the average helix length of low energy structures located by this method is only 1–3 and 1–2, respectively. The weighting factor $w = 1.0$ overemphasizes the solvation energy term, which indeed gives low values for G_{solv} (around -70 kJ/mol for 10-ALA and -120 kJ/mol for 15-ALA), but is followed by high values of E_P and negligible overall helix formation.

If, however, the balance of contributions of potential energy and solvation energy is set properly to calculate the acceptance probability $P(\Delta E_P, \Delta G_{\text{solv}})$, this cominimization of both energy terms leads to a more efficient structure optimization.

This fact is due to a higher flexibility of the peptide structure, but more importantly based on a preferential determination of pathways leading to helix formation: the peptide folding under the *pot* method gets frequently trapped in isolated minima on the potential energy hypersurface. These minima are often not preferred from the viewpoint of solvation energy, i.e., solvation decreases the acceptance probability of such dead-end situations. The overall MCSA folding process threfore shows increased dynamics, which leads preferentially to helical structures that show an *optimum* of potential energy and solvation energy, and not just a *minimum* of potential energy, as obtained under the *pot* method.[10]

Cominimization of Potential Energy and CD

The *pot_cd* method gives the cominimization of potential energy and deviation of the calculated CD spectra of every conformation generated in the course of the MCSA run with respect to the experimental spectrum [see eq. (10)]. The weighting factors 1.0, 3.0, 5.0, and 10.0 are applied, and the results of five different runs are also shown in Table II (the *pot_cd1.0*, *pot_cd3.0*, *pot_cd5.0*, and *pot_cd10.0* methods).

The most powerful setup is the *pot_cd3.0* method, where full helices are found within 10^5 MCSA steps for 10-ALA (helix lengths 8–10) as well as for 15-ALA (helix lengths 11–15). The potential energy is in the range of -4 kJ/mol for 10-ALA, which is comparable to the lowest potential energies located by the *pot* method, although the structures show helical conformation using this cominimization strategy. Obviously the "native" structure of this short peptide cannot be defined via a clear minimum on the potential energy hypersurface. Certain nonhelical structures are energetically in the same range as a fully helical conformation. This finding is in contrast to results given by Okamoto et al.,[13] who report clearly defined energetic minima for the native state of homoalanines in a length of 10.

This situation changes in the case of 15-ALA, where a full helix is energetically better separated from other structures. The lowest potential energy of a helical structure found is around -48 kJ/mol, which is far below local minima found via the *pot* method (which gives structures around 70 kJ/mol) and the *pot_solv0.5* method (showing structures around 60 kJ/mol). In both the 10-ALA and 15-ALA situation, the potential energies of helical structures are below the respective values of the ideal helix, as given in Table I. Slight variation of the ideal ϕ, ψ angles (*dihed err.* is in the range of 20°) relaxes the ideal helix to an unconstrained helix with lower potential energy.

Analyzing the cominimization results of potential energy and deviation of CD spectra shows a clear optimum for the weighting parameter w: a setting around 3.0 gives the best optimization results, and a decrease in optimization performance shows up when decreasing this weighting to 1.0 or increasing to 5.0 or 10.0. Low-weighted cominimization is still determined by the potential energies involved in computing the acceptance probability $P(\Delta E_P, \Delta CD_{\text{err}})$. High-weighted cominimization, in contrast, overemphasizes the impact of the CD spec-

trum on the structure optimization, which results in a decrease of helix length as well as an increase in potential energy.

The main problem under the optimization *pot* and *pot_solv* methods is based on the fact that the native, i.e., helical structure is neither uniquely characterized by its potential energy nor by its free energy of solvation, at least in the respective models used (force field and continuum approximation of solvation). Various random coils (mainly composed of backfolded short helices, see end: end distance distribution in Table II) show competitive energetic parameters. The CD spectrum, in contrast, reflects the overall structural characteristics of the native conformation, and the MCSA folding pathway towards structures with low energies *and* low CD_{err} is well defined. Table III reflects this situation.

Table III gives the mean energies and conformational parameters of structures accepted within MCSA step 5×10^4 and 10^5, again comparing the different optimization procedures. The *pot_cd3.0* method shows the lowest overall potential energy for both 10-ALA and 15-ALA, as well as the highest overall helix length of 7.96 for 10-ALA and 13.1 for 15-ALA for structures accepted within the second half of the MCSA run.

In contrast, the *pot* method gives a mean potential energy of 64.07 kJ/mol and an overall helix length of 5.0 for 10-ALA, the respective values for 15-ALA being 166.53 kJ/mol with a helix length of 2.4. In the course of optimization under the *pot* and *pot_solv0.5* methods, structures with reasonable helix lengths do occur (see Table II where the lowest potential energy is −5.71 kJ/mol with a helix length of 8 for 10-ALA is found), but they do not necessarily serve as an attractor towards continuous increase of helix length. Other nonhelical conformations with competitive potential and solvation energies interfere with the MCSA fold towards the full helix.

Figure 3 shows the evolution of potential energy (A), solvation energy (B) and helix length for the 15-ALA system in the course of 10^5 MCSA steps

TABLE III.

Mean Energies in kJ/mol and Conformational Parameters (out of Five Individual MCSA Runs per Method for the MCSA Step Interval $5 \cdot 10^4$–$1 \cdot 10^5$) for the 10-ALA and the 15-ALA Homopeptide Located by the Different Optimization Methods.

System	*Method*	*Mean E_P*	*Mean G_{solv}*	*Mean no. h. d.*	*Mean no. cons. h. d.*	*Mean dihed. err.*	*Mean end : end dist.*
10-ALA	*pot*	64.07	−53.01	5.80	5.02	43.33	13.88
10-ALA	*pot_solv0.5*	20.08	−40.37	7.94	7.89	28.26	15.01
10-ALA	*pot_solv1.0*	147.27	−109.35	2.04	1.03	82.70	18.90
10-ALA	*pot_cd1.0*	67.11	−59.22	4.18	3.16	48.08	14.65
10-ALA	*pot_cd3.0*	*37.31*	*−40.70*	*7.98*	*7.96*	*28.55*	*15.04*
10-ALA	*pot_cd5.0*	143.22	−79.03	2.74	1.79	68.59	14.72
10-ALA	*pot_cd10.0*	185.79	−92.69	1.30	1.12	82.61	18.30
15-ALA	*pot*	166.53	−66.71	4.10	2.46	73.46	8.73
15-ALA	*pot_solv0.5*	124.14	−82.97	6.63	3.36	58.34	15.12
15-ALA	*pot_solv1.0*	210.38	−143.62	3.41	1.03	80.38	23.24
15-ALA	*pot_cd1.0*	134.64	−79.75	5.07	2.01	67.89	3.87
15-ALA	*pot_cd3.0*	*−10.49*	*−28.51*	*13.34*	*13.19*	*20.94*	*21.61*
15-ALA	*pot_cd5.0*	104.75	−62.61	10.21	7.34	40.02	17.81
15-ALA	*pot_cd10.0*	247.16	−109.77	3.05	1.65	77.13	19.50

System denotes the 10-ALA and 15-ALA homopeptide, respectively. *Method* gives the type of optimization: *pot* denotes minimization of the potential energy only; *pot_solv* denotes cominimization of potential energy and free energy of solvation [see eq. (2)] applying the weighting factors 0.5 (*pot_solv0.5*) and 1.0 (*pot_solv1.0*); *pot_cd* means cominimization of potential energy and the deviation of experimentally determined and computed CD spectrum [see eq. (10)], the weighting factors 1.0 (*pot_cd1.0*), 3.0 (*pot_cd3.0*), 5.0 (*pot_cd5.0*), and 10.0 (*pot_cd10.0*) are applied.

Mean E_P gives the mean potential energy in kJ/mol of a structure within the particular MCSA interval; *mean G_{solv}* denotes the mean free energy of solvation in kJ/mol; *mean no.h.d.* gives the mean number of ϕ, ψ dihedral pairs within the range defined for an α-helix; *mean no.cons.h.d.* gives the mean number of consecutive ϕ, ψ dihedral pairs within the range defined for an α-helix (i.e., the helix length); *mean dihed.err.* shows the mean deviation of the backbone ϕ, ψ dihedrals from their ideal α-helix values; *mean end : end dist.* gives the mean distance between the terminal groups **N**H_2 and **C**OOH in Å of all structures accepted in the MCSA interval of interest (i.e., 5×10^4 to 10^5).

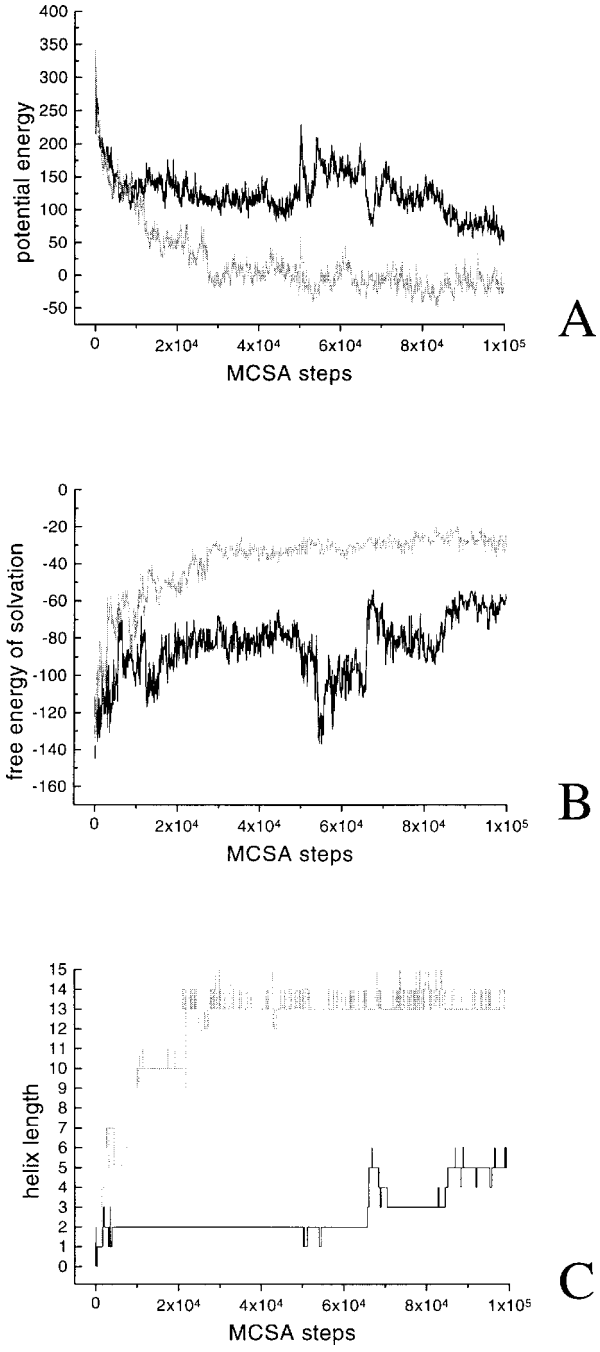

FIGURE 3. Typical evolution of the potential energy E_P (A), the free energy of solvation G_{solv} (B), and the helix length (i.e., maximum number of consecutive ϕ, ψ dihedral pairs within the angle range defined for an α-helix) (C), plotted vs. MCSA steps applying different optimization strategies on the 15-ALA homopeptide. The black line denotes the progress of the three parameters cominimizing the potential energy as well as the free energy of solvation [see eq. (2)] with the weighting factor $w = 0.5$ (*pot_solv0.5*); the gray line gives the progress of helix formation under cominimization of potential energy and CD_{err} (*pot_cd3.0*), as given in eq. (10).

comparing the *pot_solv0.5* and *pot_cd3.0* methods. Clearly, the *pot_cd3.0* method gives structures with lowest potential energy as well as with longest helix length, and consecutively with highest free energy of solvation (as the helix represents a compact structure). Once the native structure, characterized by an optimum of E_p and CD_{err}, is reached via *pot_cd3.0*, this structure is also stable during the rest of the MCSA run. All accepted structures after MCSA step 3×10^4 are within a well-defined energetic and structural range around the fully helical structure (helix lengths of 12–15). These results show a comparable efficiency as multicanonical algorithms, which locate fully α-helical structures of a length of 15 homoalanine peptides within 2×10^5 update steps.[13]

The other two methods (as well as the *pot* method, respective data not shown in Fig. 3) do, as already noted in Tables II and III, give higher potential energies as well as shorter helices. Even if a certain helix length is reached, this structural feature breaks down easily, and the fold continues towards other random coil structures.

Conclusions

Monte Carlo optimization including the Metropolis Criterion is well suited to search the conformational space of molecular systems that are characterized by a high number of degrees of freedom. However, on highly rugged potential energy hypersurfaces, as found for the peptide systems presented in this article, the optimization solely following changes in potential energy is inefficient. A large number of local minima, which are from the energetic but not from the structural viewpoint, close to the native structure, hinder an efficient search.

The consideration of a second energetic term, the free energy of solvation, in a modified Metropolis Criterion filters a large number of "dead-end" situations given by the potential energy surface. This additional information of solvation effects emphasizes folding pathways towards the native structure. The cominimization approach of potential energy and free energy of solvation, computed via a continuum approximation, increases the efficiency of the MCSA optimization algorithm. However, both methods are not capable of locating a low energy minimum close to the native structure for the test peptides 10-ALA and 15-ALA when performing only 10^5 MCSA steps.

The third approach presented in this article also performs structure optimization on the basis of the

potential energy hypersurface, but uses, in addition, a parameter representing an overall feature of the native peptide structure, the Circular Dichroism. The native peptide structure is well represented via a combination of both its potential energy and its circular dichroism spectrum, and a cominimization of the energetic situation as well as of the deviation of calculated and experimentally determined CD spectrum locates such uniquely defined structures with high efficiency: Less than 10^5 MCSA steps are necessary to fold the model peptide of a length of 15 to its final, helical state. "Dead-end" folds with low potential energy are efficiently filtered using this cominimization approach.

The CD spectrum of a peptide is usually experimentally accessible and the computed spectra reproduce well the experimental data. As long as the CD signals of the peptide of interest are strong (i.e., well-defined secondary structure elements characterize the native structure), this cominimization method may be applied.

The computation of the CD signals at every MCSA step is, furthermore, computationally efficient, and at least one order of magnitude faster than applying additional gradient minimization at every MCSA step, a procedure commonly performed to increase the performance of the MCSA optimization.[4]

We are in the process of extending this new type of circular dichroic constrained structure optimization to larger and heterogeneous systems, including combinations of helical and nonregular folds. Such structures should still be uniquely defined by their energy and the respective CD signals, and therefore, provide a good target for this new optimization method as a low potential energy, and a correct CD signal will be restricted to only a small number of structures.

References

1. Crescenzi, P.; Goldman, D.; Papadimitriou, C.; Piccolboni, A.; Yannakakis, M. J Comput Biol 1998, 5, 423; Berger, B.; Leighton, T. J Comput Biol 1998; 5, 27; Ngo, J. T.; Marks, J. Protein Eng 1992, 5, 313; Lathrop, R. H. Protein Eng 1994, 7, 1059.
2. Karplus, M.; Petsko, G. A. Nature 1990 347, 631.
3. Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. B.; Teller, A. H. J Phys Chem 1953, 21, 1087.
4. Li, Z.; Scheraga, H. A. Proc Natl Acad Sci USA 1987, 84, 6611.
5. Mayer, B.; Klein, Ch. Th.; Topchieva, I.; Köhler, G. J Comput Aid Mol Des 1999, 13, 373.
6. Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. Science 1983, 220, 671.
7. Klein, Ch. Th.; Mayer, B.; Köhler, G.; Wolschann, P. J Mol Struct (Theochem) 1996, 370, 33.
8. Wilson, S. R.; Cui, W. Biopolymers 1990, 29, 225.
9. Ma, J.; Straub, J. E. J Chem Phys 1994, 101, 533.
10. Klein, Ch. Th.; Mayer, B.; Köhler, G.; Wolschann, P. J Comput Chem 1998, 19, 1470.
11. Judson, R. S.; Jaeger, E. P.; Treasurywala, A. M.; Peterson, M. J Comput Chem 1993, 14, 1407.
12. Lee, H.; Scheraga, H. A.; Rackovsky, S. J Comput Chem 1997, 18, 1222.
13. Hansmann, U. H. E.; Okamoto, Y. J Comput Chem 1993, 14, 1333.
14. Holzwarth, G. M.; Doty, P. J Am Chem Soc 1965, 87, 218.
15. Greenfield, N.; Fasman, G. D. Biochemistry 1969, 8, 4108.
16. Woody, R. W. J Polym Sci Macromol Rev 1977, 12, 181.
17. Provencher, S. W.; Glöckner, J. Biochemistry 1981, 20, 33; Galat, A. Eur J Biochem 1996, 236, 428.
18. Tinoco, I. Adv Chem Phys 1962, 4, 113.
19. Vila, J.; Williams, R. L.; Vasques, M.; Scheraga, H. Proteins 1988, 10, 199; Eisenberg, D.; McLachlan, A. D. Nature 1986, 319, 199; Ooi, T.; Oobatake, M.; Nemethy, G.; Scheraga, H. A. Proc Natl Acad Sci USA 1987, 84, 3086.
20. Okamoto, Y. Proteins 1994, 19, 14.
21. Tobias, D. J.; Brooks, C. H., III Biochemistry 1991, 30, 6059.
22. Creamer, T. P.; Rose, G. D. Protein Sci 1995, 4, 1305.
23. Young, W. S.; Brooks, Ch. L., III J Mol Biol 1996, 259, 560.
24. Aleman, C.; Roca, R.; Luque, F. J.; Orozco, M. Proteins 1997, 28, 83.
25. Hansmann, U. H. E.; Okamoto, Y. J Chem Phys 1999, 111, 1339.
26. Samuelson, S.; Martyna, G. J. J Phys Chem B 1999, 103, 1752.
27. Padmanabhan, S.; Marqusee, S.; Ridgeway, T.; Laue, T. L.; Baldwin, R. L. Nature 1990, 344, 268; Marqusee, S.; Robins, V. H.; Baldwin, R. L. Proc Natl Acad Sci USA 1989, 86, 5286; Baldwin, R. L. Biophys Chem 1995, 55, 127; Rohl, C. A.; Fiori, W.; Baldwin, R. L. Proc Natl Acad Sci USA 1999, 96, 3682.
28. Momany, A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. J Phys Chem 1975, 79, 2361; Nemethy, G.; Pottle, M. S.; Scheraga, H. A. J Phys Chem 1983, 87, 1883; Sippl, M.; Nemethy, G.; Scheraga, H. A. J Phys Chem 1984, 88, 6231; Nemethy, G.; Gibson, K. D.; Palmer, K. A.; Yoon, C. N.; Paterlini, G.; Zagari, A.; Rumsey, S.; Scheraga, H. A. J Phys Chem 1992, 96, 6472.
29. Mayer, B.; Marconi, G.; Klein, Ch. Th.; Köhler, G.; Wolschann, P. J Incl Phenom Mol Rec Chem 1997, 29, 79.
30. Perrot, G.; Cheng, B.; Gibson, K. D.; Vila, J.; Palmer, A. N.; Nayeem, A.; Maigret, B.; Scheraga, H. A. J Comput Chem 1992, 13, 1.
31. Wesson, L.; Eisenberg, D. Protein Sci 1992, 227, 1.
32. Woody, R. W. In Circular Dichroism and the Conformational Analysis of Proteins; Fasman, G. D., Ed.; Plenum Press: New York, 1996, p. 25.
33. Schellman, J. A. Acc Chem Res 1968, 1, 144.
34. Woody, R. W. J Chem Phys 1968, 49, 4797.
35. Woody, R. W.; Tinoco, I. J Chem Phys 1967, 46, 4927.

36. Obbink, J. H.; Hezemans, A. M. F. Theor Chim Acta 1976, 43, 74.
37. Quadrifoglio, F.; Urry, D. W. J Am Chem Soc 1968, 90, 2755.
38. Chen, Y.-H.; Yang, J. T.; Chau, K. H. Biochemistry 1974, 13, 3350.
39. Zimm, B. H.; Bragg, K. J. J Chem Phys 1959, 31, 526.